Biological Applications of Knowledge Graph Embedding Models

Sameh K. Mohamed [1,2*], Vít Nováček [1,2] and Aayah Nounu [3]

Published in the Briefings in Bioinformatics journal on 17 February 20209

Abstract

Complex biological systems are traditionally modelled as graphs of interconnected biological entities. These graphs, *i.e.* biological knowledge graphs, are then processed using graph exploratory approaches to perform different types of analytical and predictive tasks. Despite the high predictive accuracy of these approaches, they have limited scalability due to their dependency on time-consuming path exploratory procedures. In recent years, owing to the rapid advances of computational technologies, new approaches for modelling graphs and mining them with high accuracy and scalability have emerged. These approaches, *i.e.* knowledge graph embedding models, operate by learning low-rank vector representations of graph nodes and edges that preserve the graph's inherent structure. These approaches were used to analyse knowledge graphs from different domains where they showed superior performance and accuracy compared to previous graph exploratory approaches. In this work, we study this class of models in the context of biological knowledge graphs and their different applications. We then show how knowledge graph embedding models can be a natural fit for representing complex biological knowledge modelled as graphs. We also discuss their predictive and analytical capabilities in different biology applications. In this regard, we present two example case studies that demonstrate the capabilities of knowledge graph embedding models: prediction of drug target interactions and polypharmacy side-effects. Finally, we analyse different practical considerations for knowledge graph embeddings, and we discuss possible opportunities and challenges related to adopting them for modelling biological systems.

Key words: biomedical knowledge graphs, knowledge graph embeddings, tensor factorisation, link prediction, drug target interactions, polypharmacy side-effects.

[1] Data Science Institute, College of Engineering and Informatics, NUI Galway, Galway, Ireland
[2] Insight Centre for Data analytics, NUI Galway, Galway, Ireland
[3] MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK
*To whom correspondence should be addressed.

1 Introduction

Biological systems consist of complex interconnected biological entities that work together to sustain life in living systems. This occurs through complex and systematic biological interactions of the different biological entities. Understanding these interactions is key to elucidating the mechanism-of-action of the different biological functions (*e.g.* angiogenesis, metabolism, apoptosis, etc), and thus, understanding causes and activities of diseases and their possible therapies. This encouraged the development of multiple physical and computational methods to assess, verify and infer different types of these interactions. In this study, we focus on the use of computational methods for assessing and inferring interactions (associations) between different biological entities at the molecular level. We hereof study the use of knowledge graphs and their embedding models for modelling molecular biological systems and the interactions of their entities.

Initially, basic networks *i.e.* uni-relational graphs, were adopted by early efforts for modelling complex interactions in biological systems [1–4]. Despite their initial success [5], these networks could not preserve the semantics of different types of associations between entities. For example, protein-protein interaction networks modelled with basic networks cannot differentiate between different types of interactions such as inhibition, activation, phosphorylation, etc. Therefore, more recent works modelled biological systems using heterogeneous multi-relational networks *i.e.* knowledge graphs, where they utilised different visual [6, 7] and latent representations [8, 9] of graph entities to infer associations between them.

In the context of biological applications, knowledge graphs were used to model biological data in different projects such as the UNIPROT [10], Gene Ontology [11] and Bio2RDF [12] knowledge bases. Moreover, they were the basis of multiple predictive models for drug adverse reactions [6, 8], drug repurposing [9, 13] and other predictions for different types of biological concepts associations [13, 14]. The task of learning biological associations in this context is modelled as link prediction in knowledge graphs [15]. Predictive models then try to infer a typed link between two nodes in the graph using two different types of features: graph features and latent-space vector representations.

Graph features models (*i.e.* visual feature models) are part of the network analysis methods which learn their predictions using different feature types such as random walks [16, 17], network similarity [18], nodes connecting paths [19] and subgraph paths [19, 20]. They are used in multiple biological predictive applications such as predicting drug targets [21] and protein–protein interaction analysis [18]. Despite the expressiveness of graph feature models predictions, they suffer from two major drawbacks: limited scalability and low accuracy [22, 23]. They are also focused on graph local features compared to embedding models which learn global latent features of the processed graph.

Latent feature models *i.e.* embedding models, on the other hand, express knowledge graphs' entities and relations using low-rank vector representations that preserve the graph's global structure. Knowledge graph embedding (KGE) models on the contrary are known to outperform other approaches in terms of both the accuracy and scalability of their predictions despite their lack of expressiveness [23–25].

In recent years, knowledge graph embedding models witnessed rapid developments that allowed them to excel in the task of link prediction [24–30]. They have then been widely used in various applications including computational biology in tasks like predicting drug target interactions [9] and predicting drug polypharmacy side-effects [8]. Despite their high accuracy predictions in different biological inference tasks, knowledge graph embeddings are in their early adoption stages in computational biology. Moreover, many computational biology studies that have used knowledge graph embedding models adopted old versions of these models [31, 32]. These versions have then received significant modifications through recent computer science research advances [25].

In a previous study, Su et. al. [14] have introduced the use of network embedding methods in biomedical data science. The study compiles a taxonomy of embedding methods for both basic and heterogeneous networks where it discusses a broad range of potential applications and limitation. The study's objective was to introduce the broad range of network embedding methods, however, it lacked deeper investigation into the technical capabilities of the models and how can they be integrated with a specific biological problem. The study also did not compare the investigated models in terms of their accuracy and scalability which is essential to assist reader from the biological domain to understand the key differences between these methods as to their applicability.

In this study, we exclusively explore KGE models, focusing on the best performing models in terms of both scalability and accuracy across various biological tasks. We use these case studies to demonstrate the analytical capabilities of KGE models, *e.g.* learning clusters and similarity measures in different biological problems. We also explore the process of building biological knowledge graphs for generic and specific biological inference tasks. We then present computer-based experimental evaluation of knowledge graph embedding models on different tasks such as predicting drug target interactions, drug polypharmacy side-effects and prediction of tissue-specific protein functions.

The rest of this study is organised as follows: Sec. 2.1 discusses knowledge graphs as a data modelling technique and their applications in the biological domain. Sec. 2.2 discusses knowledge graph embedding (KGE) models, their design and how they operate on different types of data. Sec. 3 presents the example case studies that we will use throughout the study. Sec. 4 discusses the predictive and analytical capabilities of KGE models on the designated case studies discussed in Sec. 3. Sec. 5 discusses the performance of KGE models on biological data in terms of the predictive accuracy and scalability. Sec. 6 discusses the current challenges and possible opportunities of the use of KGE models to model the different types of biological systems. Finally, we discuss our conclusions in Sec. 7.

2 Background

In this section, we discuss both knowledge graphs and knowledge graph embedding models in the context of biological applications.

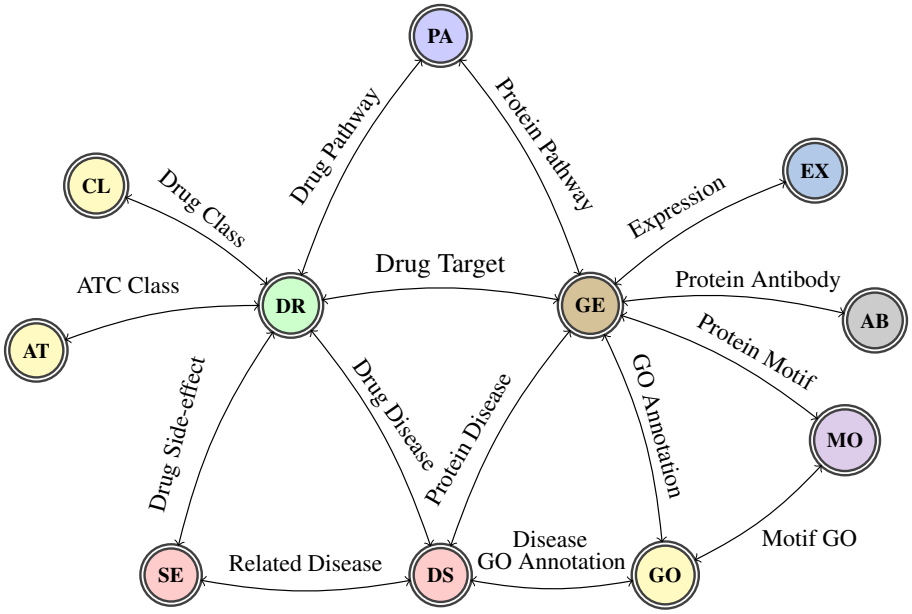

Figure 1: A schema of a knowledge graph that models a complex biological system of different types of entities and concepts. The abbreviation DR represents drugs, GE represents proteins (their genes), EX represents protein expressions (tissues and cell-lines), AB represents protein antibodies, MO represents protein motifs and other sequence annotations, GO represents gene ontology, DS represents diseases, SE represents drug side-effects, AT represents ATC classes, CL represents drug classes and PA represents pathways.

2.1 Knowledge graphs

A knowledge graph is a data modelling technique that models linked data as a graph, where the graph's nodes represent data entities and its edges represent the relations between these entities. In recent years, knowledge graphs became a popular means for modelling relational data where they were adopted in various industrial and academic applications such as semantic search engines [33], question answering systems [34] and general knowledge repositories [35]. They were also used to model data from different types of domains such as general human knowledge [35], lexical information [36] and biological systems [12].

Knowledge graphs model facts as subject, predicate and object (SPO) triples, where subjects and objects are the knowledge entities and predicates are the knowledge relations. In this context, the subject entity is associated to the object entity with the predicate relation *e.g.* (*Aspirin*, *drug_target*, *COX1*). Fig. 1 shows an illustration of a schema of a knowledge graph that models complex associations between different types of biological entities such as drugs, proteins, antibodies, etc. It also models different types of relations between these entities, where these relation carry different

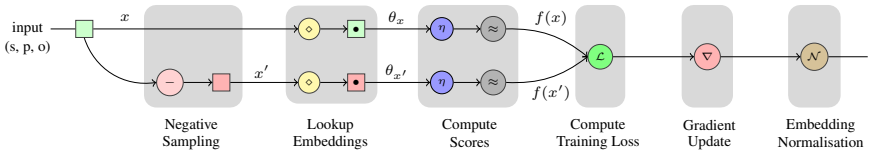

Figure 2: An illustration of the training network of one training instance of a knowledge graph embedding model.

association semantics.

In our study, we use $\mathscr{G}$ to denote a knowledge graph, $\mathscr{E}$ to denote entities and $\mathscr{R}$ to denote relations *i.e.* predicates. We also use $\mathscr{N}_e$ and $\mathscr{N}_r$ to denote the total count of both entities and relations in a knowledge graph respectively.

Popular Biological Sources. Online knowledge bases are a popular means for publishing large volumes of biological data [37]. In recent years, the number of these knowledge bases have grown, where they cover different types of data such as paper abstracts [38], raw experimental data [39], curated annotations [10, 40, 41], etc. Biological knowledge bases store data in different structured and unstructured (free text *e.g.* comments) forms. Although both data forms can be easily comprehended by humans, structured data is significantly easier for automated systems. In the following, we explore popular examples of these knowledge bases which offer structured data that can be easily and automatically consumed to generate knowledge graphs.

Table 1 summarises the specialisations and the different types of covered biological entities of a set of popular biological knowledge bases. The table also shows that most of the current knowledge bases are compiled around proteins (genes). However, it also shows their wide coverage of the different types of biological entities such as drugs, their indications, gene ontology annotations, etc.

Building Biological Knowledge Graphs. Knowledge graphs store information in a triplet form, where each triplet (*i.e.* triple) model a labelled association between two unique unambiguous entities. Data in biological knowledge bases, however, lacks these association labels. Different knowledge bases also use different identifier systems for the same entity types which results in the ambiguity of entities of merged databases. Building biological knowledge graph process therefore mainly deals with these two issues.

In the association labelling routine, one can use different techniques to provide meaningful labels for links between different biological entities. This, however, is commonly achieved by using entity types of both subject and object entities to denote the relation labels as shown in Fig 1 (*e.g.* "Drug Side-effect" as a label for link between two entities that are known to be types of Drug and Side-effect, respectively).

The ambiguity issue, *i.e.* merging entities of different identifier systems, is commonly resolved using identifier mapping resource files. Different systems study entities on different speciality levels. As a result, the links between their different identifier systems is not always in a form of one-to-one relationships. In such cases, a

	Properties		Entity coverage								
Knowledge base	**Format**	**Speciality**	**Proteins**	**Drugs**	**Indications**	**Diseases**	**Gene Ontology**	**Expressions**	**Antibodies**	**Phenotypes**	**Pathways**
UNIPROT [10]	S/U	GE	✓	✓		✓	✓	✓	✓		✓
REACTOME [42]	S	PA	✓				✓				✓
KEGG [40, 43]	S	PA	✓	✓		✓					✓
DrugBank [44]	S/U	DR	✓	✓							✓
Gene Ontology [11]	S	GO	✓				✓				✓
CTD [45]	S/U	CH	✓	✓			✓			✓	✓
ChEMBL [46]	S/U	CH	✓	✓	✓	✓		✓			
SIDER [47]	S	DR		✓	✓						
HPA [48]	S/U	GE	✓				✓	✓	✓		
STRING [49]	S	GE	✓								
BIOGRID [50]	S	GE	✓								
InAct [41]	S	GE	✓								
InterPro [51]	S	GE	✓								
PharmaGKB [52]	S	DR	✓	✓							
TTD [53]	S	DR	✓	✓							
Supertarget [54]	S	DR	✓	✓							

Table 1: A comparison between popular biological knowledge graph in terms of the coverage of different types of biological entities. The abbreviation S represent structured data, U represents unstructured data, DR represents drugs, GE represents proteins, GO represents gene ontology,PA represents pathways and *CH* denotes chemicals.

decision is made to apply a specific filtering strategy based on either expert's opinion or problem-specific properties (for instance, deciding on an authoritative resource such as UniProt for protein entities and resolving all conflicts by sticking to that resource's naming scheme and conventions).

To complement the basic principles introduced in the previous paragraphs, we refer the reader to the Bio2RDF initiative [55] that has extensively studied the general topic of building interlinked biological knowledge graphs (see also Bio2RDF scripts[1] for corresponding scripts and conversion convention details). General principles as well as an example of actual implementation of conversion from (relational) databases into RDF (*i.e.* knowledge graphs) are discussed in the study of Bizer et. al. [56]. Possible solutions to the problem of aligning and/or merging several such knowledge graphs are reviewed in the study of Amrouch et.al. [57] that focuses on ontology matching. An example of a more data-oriented method is for instance LIMES [58]. All these approaches may provide a wealth of inspiration for building

[1] https://github.com/bio2rdf/bio2rdf-scripts/wiki

bespoke approaches to building knowledge graphs in specific biomedical use cases, should the information we provide in this section be insufficient.

2.2 Knowledge graph embeddings (KGE)

In this section, we discuss knowledge graph embedding models where we briefly explore their learning procedure. We then explore different embedding representation types and their potential uses and application.

The learning procedure. Multiple studies have explored knowledge graph embedding (KGE) models, their technical design, training objectives and predictive capabilities on general benchmarking settings [15, 24, 59]. Therefore, in the following we only focus on providing a brief and concise description of how KGE models work.

KGE models operate by learning low-rank representations of knowledge graph entities and relations. The KGE learning step is a multi-phase procedure as shown in Fig. '2 which is executed iteratively on knowledge graph data. Initially, all entities and relations are assigned random embeddings (noise). They are then updated using a multiphase learning procedure.

KGE models consume knowledge graphs in the form of subject, predicate and object (spo) triplets. They first generate negative samples from the input true triplets using uniform random corruptions of the subjects and objects [60]. KGE models then lookup corresponding embedding of both the true and corrupted triplets. The embeddings are then processed using model-dependent scoring functions (cf. mechanism-of-action in Table 2) to generate scores for all the triplets. The training loss is then computed using model-dependent loss functions where the objective is to maximise the scores of true triplets and minimise the scores of corrupted triplets. This objective can be formulated as follows:

$$\forall_{t\in\mathbb{T},t'\in\mathbb{T}'} f(\theta_t) > f(\theta_{t'}), \tag{1}$$

where $\mathbb{T}$ denotes the set of true triplets, $\mathbb{T}'$ denotes the set of corrupted triplets, f denotes the model-dependent scoring function and θ_t denotes the embeddings of the triplet t.

Traditionally, KGE models use a ranking loss, *e.g.* hinge loss or logistic loss, to model the objective training cost [26, 28, 29]. This strategy allows KGE models to efficiently train their embeddings in linear time, $\mathcal{O}(d)$, where K denotes the size of the embedding vectors. On the other hand, some KGE models such as the ConvE [30] and the ComplEx-N3 [25] models adopt multi-class based strategies to model their training loss. These approaches have shown superior predictive accuracy compared to traditional ranking based loss strategies [25, 30]. However, they suffer from limited scalability as they operate on the full entity vocabulary.

The KGE models minimise their training loss using different variations of the gradient descent algorithm *e.g.* Adagrad, AMSGrad, etc. Finally, some KGE models normalise their embeddings as a regularisation strategy to enhance their generalisation. This strategy is often associated to models which adopt ranking based training

Model	Scoring mechanism	Em. Format	Time	Space	Year	Repository (Github)
RESCAL [27]	Tensor factorisation	(d, d^2)	$\mathcal{O}(d^2)$	$\mathcal{O}(nd + md^2)$	2011	mnick/rescal.py
TransE [26]	Linear translation	(d, d)	$\mathcal{O}(d)$	$\mathcal{O}(nd + md)$	2014	ttrouill/complex
DistMult [28]	Bilinear dot product	(d, d)	$\mathcal{O}(d)$	$\mathcal{O}(nd + md)$	2015	ttrouill/complex
HolE [62]	Fast Fourier Transformation	(d, d)	$\mathcal{O}(d \log d)$	$\mathcal{O}(nd + md)$	2016	mnick/holographic-embeddings
ComplEx [29]	Complex product	$(2d, 2d)$	$\mathcal{O}(d)$	$\mathcal{O}(nd + md)$	2016	ttrouill/complex
ANALOGY [63]	Analogical structure	(d, d)	$\mathcal{O}(d)$	$\mathcal{O}(nd + md)$	2017	quark0/ANALOGY
ConvE [30]	Convolutional filters	(d, d)	$\mathcal{O}(d)$	$\mathcal{O}(nd + md)$	2018	TimDettmers/ConvE
TriModel [64]	Multi-part embeddings	$(3d, 3d)$	$\mathcal{O}(d)$	$\mathcal{O}(nd + md)$	2019	samehkamaleldin/libkge

Table 2: A comparison between popular KGE models, their learning mechanism, published year and available code bases. Em. format column denotes the format of the model embeddings in the form $(g(d), h(d))$, where d denotes the embeddings size, $g(d)$ denotes the shape of the entities embeddings and $h(d)$ denotes the shape of the relations embeddings. n and m denote the number of entities and relations respectively in the space complexity column.

loss strategies such as the TransE and DistMult models [26, 28].

The learning multi-phase procedure is executed iteratively to update the model's embeddings until they reach an optimal state that satisfies the condition in Eq. 1. Table 2 also provides a summary of properties of popular KGE models, their mechanism of action *i.e.* scoring mechanism, output embeddings format, runtime complexity, release year and available code bases.

Knowledge graph embedding models ingest graph data in triplets form where they learn global graph low-rank latent features which preserve the graph's coherent structure. These features encode semantics such as node types and their neighbours by isolating nodes' embeddings on different embedding dimensions [23]. However, they have limited ability to encode indirect semantics such as logical rules and indirect relations [61].

Embedding representation. Knowledge graph embeddings have different formats *e.g.* vectors, matrices, etc, which serve as numerical feature representations of their respective objects. These representations can be used in both general tasks such as clustering and similarity analysis, as well as in specific inference tasks such as predicting different association types. Similarly, in computational biology, they can be used to cluster biological entities such as protein, drugs, etc, as well as to learn specific biological associations such as drug targets, gene related diseases, etc. Embeddings of biological entities can also be used as representative features in traditional regression and classification models *e.g.* logistic regression or SVM classifiers.

Popular KGE models. Table 2 presents a comparison between a set of popular KGE models, their scoring mechanism, embeddings format, time complexity, space complexity, year of publication, and corresponding source code repository. These models use different approaches to learn their embeddings where they can be categorised into three categories: distance based models, factorisation based models and convolutional models. Distance based models such as the TransE model use linear translations to model their embeddings interactions using a linear time and space complexity procedure. Convolution based methods such as the ConvE use

convolutional neural networks to model embedding interactions which also have a linear time and space complexity. Factorisation based models, on the other hand, use dot product based procedures to model embedding interactions, where they also have linear time and space complexity. However, tensor factorisation based models commonly use higher rank embeddings than convolution and distance based models [29, 64].

In this study, we are focused on embedding methods which operate on multi-relational graphs as we mentioned in the introduction of the paper. The DeepWalk [65], Node2Vec [66], etc are uni-relational graphs embedding methods, thus, they we do not include them in this study.

3 Examples of biological case studies

In the following, we present two example biological case studies that we use through this study to demonstrate the capabilities of KGE models. Firstly, we discuss the task of predicting drug target interactions where we model biological information as a knowledge graph. We then evaluate the predictive accuracy of KGE models and we compare them to other state-of-the-art approaches. Secondly, we discuss the task of predicting drug polypharmacy side-effects, where we model the investigated drug polypharmacy data as a 3D tensor.

3.1 Predicting drug target interactions

The study of drug targets has become very popular with the objective of explaining mechanisms of actions of current drugs and their possible unknown off-target activities. Knowing targets of potential clinical significance also plays a crucial role in the process of rational drug development. With such knowledge, one can design candidate compounds targeting specific proteins to achieve intended therapeutic effects. Large-scale and reliable prediction of drug-target interactions (DTIs) can substantially facilitate development of such new treatments. Various DTI prediction methods have been proposed to date. Examples include chemical genetic [67] and proteomic methods [68] such as affinity chromatography and expression cloning approaches. These, however, can only process a limited number of possible drugs and targets due to the dependency on laboratory experiments and available physical resources. Computational prediction approaches have therefore received a lot of attention lately as they can lead to much faster assessments of possible drug-target interactions [69, 70].

Data. We consider the DrugBank_FDA [71] benchmarking dataset as an example to evaluate the predictive accuracy of KGE models and to compare them to other approaches. We also utilise the UNIPROT [10] database to provide richer information about both drugs and their protein targets in the input knowledge graph. The dataset contains 9881 known drug target interactions which involve 1482 drugs and 1408 protein targets.

Related work. The work of Yamanishi et. al.[69] was one of the first approaches to

predict drug targets computationally. Their approach utilised a statistical model that infers drug targets based on a bipartite graph of both chemical and genomic information. The BLM-NII [70] model was developed to improve the previous approach by using neighbour-based interaction-profile inference for both drugs and targets. More recently, Cheng et. al. [72, 73] proposed a new way for predicting DTIs, where they have used a combination of drug similarity, target similarity and network-based inference. The COSINE [74] and NRLMF [75] models introduced the exclusive use of drug-drug and target-target similarity measures to infer possible drug targets. This has an advantage of being able to compute predictions even for drugs and targets with limited information about their interaction data. However, these methods only utilised a single measure to model components similarity. Other approaches such as the KronRLS-MKL [76] model used a linear combination of multiple similarity measures to model the overall similarity between drugs and targets. Non-linear combinations were also explored in an early study [70] and shown to provide better predictions. Recently, further predictive models were developed to utilise matrix factorisation [77] and biological graph path features [7] to enable more accurate drug target prediction.

3.2 Predicting polypharmacy side-effects

Polypharmacy side-effects are a specific case of adverse drug reactions that can cause significant clinical problems and represent a major challenge for public health and pharmaceutical industry [78]. Pharmacology profiling leads to identification of both intended (target) and unintended (off-target) drug-induced effects, i.e. biological system perturbations. While most of these effects are discovered during pre-clinical and clinical trials before a drug release on the market, some potentially serious adverse effects only become known when the drug is in use already.

When more drugs are used jointly (i.e. polypharmacy), the risk of adverse effects rises rather rapidly [79, 80]. Therefore, reliable automated predictions of such risks are highly desirable to mitigate their impact on patients.

Data. In this case study, we consider the dataset compiled by Zitnik et al. [8] as an example benchmark. The dataset includes information about multiple polypharmacy drug side-effects [2]. The dataset also contains facts about single drug side-effects, protein-protein interactions and protein-drug targets. The drug side-effects represented in the dataset are collected from the SIDER (Side Effect Resource) database [47] and the OFFSIDES and TWOSIDES databases [80]. These side-effects are categorised into two groups: mono-drug and polypharmacy drug-drug interaction side-effects.

In our study, we only consider the polypharmacy side-effects and we filter out both the mono-side effects and drug targets data.

Related work. The research into predictive approaches for learning drug polypharmacy side effects is in its early stages [8]. The decagon model [8] is one of the first introduced methods for predicting polypharmacy side-effects which models the polypharmacy side-effects data as a knowledge graph. It then solves the problem as a

[2] http://snap.stanford.edu/decagon/

link prediction problem using a generative convolution based strategy. Despite its effectiveness, this approach still suffers from a high rate of false positives. Furthermore, other approaches considered using a multi-source embedding model [81] to learn representations of drugs and polypharmacy side-effects. These approaches achieved similar performance to the Decagon model with a more scalable training procedure [81].

3.3 Predicting tissue-specific protein functions

Proteins are usually expressed in specific tissues within the body where their precise interactions and biological functions are frequently dependent on their tissue context [82, 83]. The disorder of these interactions and functions results in diseases [84, 85]. Deep understanding of tissue-specific protein activities is therefore essential to elucidate the causes of diseases and possible treatments.

Data. We consider the tissue-specific dataset compiled by Zitnik et. al [86] to study tissue-specific protein functions. The dataset contain protein-protein interactions and protein functions of 144 tissue types[3].

Related work. Recently, Zitnik et. al. have developed the state-of-the-art model, the OhmNet model [86], a hierarchy-aware unsupervised learning method for multi-layer networks. It models each tissue information as a separate network, and learns efficient representations for proteins and functions by generating their embeddings using the tissue-specific protein-protein interactome and protein functions. They have also examined other different approaches such as the LINE model [87] which uses a composite learning technique where it learns half of the embeddings' dimensions from the direct neighbour nodes, and the other half from the second hop connected neighbours. The GeneMania model [88] is another model which has suggested a propagation based approach for predicting tissue-specific protein functions. In this method, the tissue-specific networks are firstly combined into one weighted network, and they are then propagated to allow predicting other unknown protein functions.

4 Capabilities of KGE models

KGE models can be used in different supervised and unsupervised applications where they provide efficient representations of biological concepts. They can be used in applications such as learning biological associations, concepts similarity and clustering biological entities. In this section, we discuss these applications in different computational biology tasks. We provide a set of example uses cases where we present the data integrated in each example, how the KGE models were utilised and we report the predictive accuracy of the KGE models and we compare it to other approaches when possible.

[3] http://snap.stanford.edu/ohmnet/

4.1 Learning biological associations

KGE models can process data in the form of a knowledge graph. They then try to learn low-rank representations of entities and relations in the graph which preserve its coherent structure. They can also process data in a three dimensional (3D) tensor form where they learn low-rank representations for the tensor entities that preserve true entity combination instances in the tensor.

In the following, we provide two examples for learning biological associations on a knowledge graph and a 3D tensor in a biological application. First, we discuss the task of predicting drug target interactions where we model biological information as a knowledge graph. We then evaluate the predictive accuracies of KGE models and we compare them to other state-of-the-art approaches. Secondly, we discuss the task of predicting drug polypharmacy side-effects, where we model the related data as a 3D tensor. We then apply KGE models to perform tensor factorisation and we evaluate their predictive accuracy in learning new polypharmacy side-effects compared to other state-of-the-art approaches.

- **Drug target prediction benchmark** We present a comparison between state-of-the-art drug target predictors and knowledge graph embedding models in predicting drug target interactions. The KGE models in this context utilise the fact that the current drug target knowledge bases like DrugBank [71] and KEGG [40] are largely structured as networks representing information about drugs and their relationship with target proteins (or their genes), action pathways, and targeted diseases. Such data can naturally be interpreted as a knowledge graph. The task of finding new associations between drugs and their targets can then be formulated as a link prediction problem on a biological knowledge graph.

  We use the standard evaluation protocol for the drug target interaction task [7] on the DrugBank_FDA dataset that we introduced in Sec. 3.1. We use a 5-fold cross validation evaluation on the drug target interactions where they are divided into splits with uniform random sampled negative instances with a 1:10 positive to negative ratio.

  Fig. 3 presents the outcome results of the KGE models (DistMult, ComplEx and TriModel) compared to other approaches (DDR [7], DNILMF [77], NRLMF [77], NRLMF [75], KRONRLS-MKL [76], COSINE [89], and BLM-NII [70]) on the DrugBank_FDA dataset. The figure shows that the KGE models outperform all other approaches in terms of both the area under the ROC and precision recall curves.

- **Polypharmacy side-effects prediction benchmark** In Sec. 3.2 we discussed the problem of predicting polypharmacy side-effects, the currently available data and related works. In the following, we present an evaluation benchmark for present polypharmacy side-effects where we compare the KGE models with current state-of-the-art approaches. We first split the data into two sets, train and test splits, where the two splits represent 90% and 10% of the data respectively. We then generate random negative polypharmacy side-effects

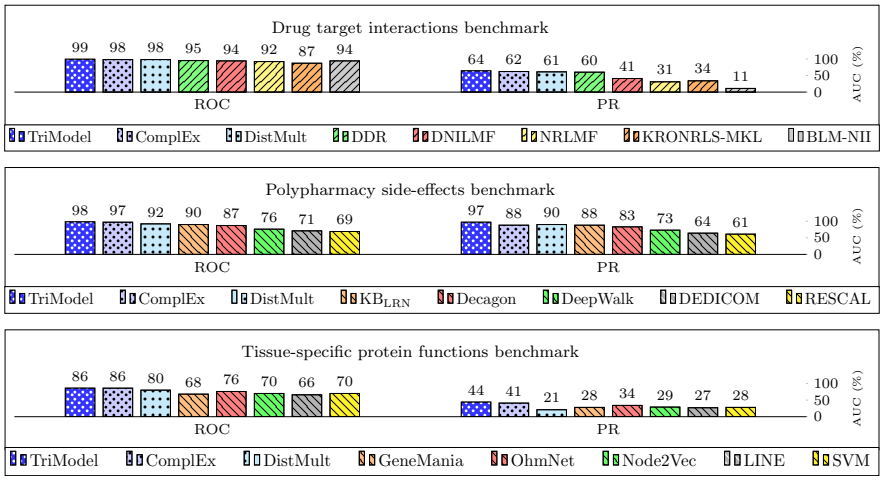

Figure 3: A summary of results of an evaluation of the predictive accuracy of knowledge graph embedding models compared to other models on two biological inference tasks: predicting drug targets and predicting polypharmacy side-effects. The reported results represent the score percentage of the area under the ROC and precision recall curves for the left and right side bars respectively.

by randomly generating combinations of drugs for each polypharmacy side effect where the ratio between negative and positive instances is 1:1. We only consider drug combinations that did not appear in the both training and test splits to enhance the quality of sampled negatives and decrease the ratio of false negatives.

We use the holdout test defined by Zitnik et. al. [8] where we train the predictive models on the training data and test their accuracy on the testing data split. We also run a 5-runs averaged 5-fold cross validation evaluation to ensure the consistency of the model reported results over the different folds, however, we only report the holdout test results which are comparable with state-of-the-art methods. Our k-fold cross validation experiments confirm that the model results are similar or insignificantly different across different random testing splits.

We use the area under the ROC and precision recall metrics to assess the quality of the predicted scores. Fig. 3 presents the results of our evaluation where we compare KGE models such as the DistMult, ComplEx and TriModel models to the current popular approaches (Decagon [8], KB_LRN [91], RESCAL [27], DEDICOM [92], DeepWalk [65]). The results show that KGE models outperform other state-of-the-art approaches in terms of both the area under the ROC and precision recall curves.

- **Tissue-specific protein function prediction benchmark** In Sec. 3.3 we have presented the problem of tissue-specific protein function prediction bench-

mark where we have discussed current predictive models and established benchmarking datasets. In the following, we present an evaluation benchmark between a set of traditional approaches such as the OhmNet [86], LINE [87], GeneMania [88] and SVM [86] models and other KGE models. We use the dataset generated by Zitnik et. al. [86] which provides training and testing data with both positive and negative instances where the negative to positive ratio is 1 to 10.

We conduct a holdout test using the provided training and testing dataset where we train our models on the training split and evaluate them on the testing using the area under the ROC and precision recall curves. Fig. 3 presents the outcome of our experiments where it shows that KGE models such as the TriModel and ComplEx models achieve the best results in terms of both the area under the ROC and precision recall curves. Similar to the previous experiments, we also ran a 5-runs 5-fold cross validation test to ensure the consistency of our results and the results of our experiments confirm the results reported in the holdout test. However, we only report the holdout test results to be able to compare to other approaches.

In all of our holdout test experiments, we learn the best hyperparameters using a grid-search on the validation data split, where the training set is divided into two sets for training and validation (90% and 10% respectively) in the absence of a validation set. On the other hand, in the cross validation experiments, we re-split each into training and validation splits (90% and 10% respectively) in order to learn the model's best hyperparameters. We have found the the embedding size is the most sensitive hyperparameters where it correlates with the graph size. The regulation weight and and embedding dropout also are important hyperparameters which affect the generality of the models from the validation to the testing split.

Example source code scripts and datasets of the experiments which we executed in this study are available at: `https://github.com/samehkamaleldin/bio-kge-apps`.

4.2 Learning similarities between biological entities

The KGE models enable a new type of similarity which can be measured between any two biological entities using the similarity between their vector representation. The similarity between vectors can be computed using different techniques such as the *cosine* and *p-norm* similarities. Since the KGE representation is trained to preserve the knowledge graph structure, the similarity between two KGE representations reflects their similarity in the original knowledge. Therefore, the similarities between vector representations of KGE models, which are trained on a biological knowledge graphs, represent the similarities between corresponding entities in the original knowledge graph.

In the following, we explore a set of examples for using KGE similarities on biological knowledge graphs. We have used the drug-target knowledge graph created for the drug target prediction task to learn embeddings of drugs, their target proteins and the entities of the motifs of these proteins according to the PFam database [90]. We have then computed the similarities between embeddings of entities of the same

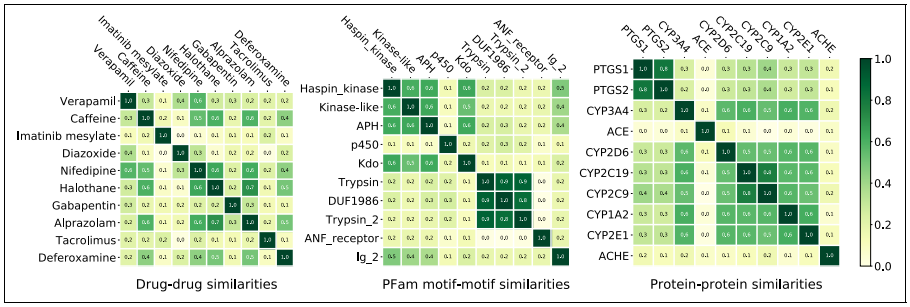

Figure 4: Three similarity matrices that denote the Drug-drug similarities, motif-motif similarities and protein-protein similarities. The similarity values are generated by computing the cosine similarity between the embeddings of the pairs of compared entities. All the embeddings used to generated this figure are computed on the DrugBank_FDA datasets with the proteins associated to their PFam [90] motifs and protein families.

type such as drugs, proteins and motifs as shown in Fig. 4. All the similarity scores in the illustration are computed using cosine similarity between the embeddings of the corresponding entity pair. The results show that the similarity scores are distributed from 0.0 to 1.0, where the 0.0 represents the least similar pairs and the 1.0 scores represent the similarity between the entity and itself. We then assess the validity of resulting scores by investigating the similarity of attributes of a set of the examined concepts with highest and lowest scores.

- **Drug-drug embedding similarity** The left similarity matrix in Fig. 4 illustrates the drug-drug similarity scores between the set of the most frequent drugs in the DrugBank_FDA dataset. The scores are computed on the embeddings of drugs learnt in the drug target interaction training pipeline. The figure shows that the majority of drug pairs have a low similarity (0.0 ~ 0.2). For example, the similarity score between the drug pairs (*Diazoxide*, *Caffeine*) and (*Tacromlimus*, *Diazoxide*) are zero. We asses these results by assessing the commonalities between the investigated drugs in terms of indications, pharmacodynamics, mechanism of action, targets, enzymes, carriers and transporters. The *Caffeine* and *Diazoxide* in this context have no commonalities except for that they are both diuretics [93, 94]. On the other hand, Halothane and Alprazolam does not share any of the investigated commonalities.

  The results also shows a few drug-drug similarities with relatively higher scores (0.6 ~ 0.7). For example, the similarity scores of the drug pairs (*Alprazolam*, *Halothane*), (*Alprazolam*, *Caffeine*) and (*Halothane*, *Caffeine*) are 0.7, 0.6 and 0.6 respectively. These finding can be supported by the fact that the two drug pairs share common attributes in terms of their targets, enzymes and carriers. For example, both *Alprazolam* and *Halothane* act on sedating individuals and they target the GABRA1 protein [95, 96]. They are also broken by CYP3A4 and CYP2C9 enzymes and carried by albumin [97]. Similarly, the (*Alprazolam*,

Caffeine) and (*Halothane*, *Caffeine*) pairs have common associated enzymes.

- **Motif-motif embedding similarity** The middle similarity matrix in Fig. 4 illustrates the motif-motif similarity scores between the set of the most frequent PFam motifs associated with protein targets from the drug target interaction benchmark. The lowest motif-motif KGE based similarity scores correspond to the pairs (*ANF_receptor*, Trypsin) , (*ANF_receptor*, *DUF1986*) and (*ANF_receptor*, *Trypsin_2*).

  On the other hand, The highest similarity scores (0.8, 0.9 and 0.9) exist between the pairs (*Trypsin*, *DUF1986*), (*Trypsin_2*, *DUF1986*) and (*Trypsin*, *Trypsin_2*) respectively.

  We assess the aforementioned findings by investigating the nature and activities of each of the discussed motifs. For example, *Trypsin* is a serine protease that breaks down proteins and cleaves peptide chains while *Trypsin_2* is an isozyme of *Trypsin* which has a different amino acid sequence but catalyses the same chemical reaction as *Trypsin* [98].

  Moreover, the DUF1986 is a domain that is found in both of these motifs which supports the high similarity scores. On the other hand, the *ANF receptor* is an atrial natriuretic factor receptor that binds to the receptor and causes the receptor to convert *GTP* to *cGMP*, and it plays a completely different role to trypsin, which supports its reported low similarity scores with trypsin.

- **Protein-protein embedding similarity** The right similarity matrix in Fig. 4 illustrates the protein-protein similarity scores between the set of the most frequent protein targets from the drug target interaction benchmark. The highest scored protein-protein pairs are (*PTGS1*, *PTGS2*) and (*CYP2C19*, *CYP2C9*) with the scores 0.8 and 0.8 respectively. This can be supported by the fact that the proteins CYP2C9, CYP1A2 and CYP2E1 belong to the same family of enzymes and thus they have similar roles.

  On the other hand, The *ACE* protein have the lowest similarity scores with the *CYP2C9*, *CYP1A2* and *CYP2E1* proteins with 0.0 similarity score. This can be supported by the fact that ACE is a a hydrolase enzyme which is completely different from *CYP2C9*, *CYP1A2* and *CYP2E1* which are Oxidoreductases enzymes.

4.3 Clustering biological entities

In the following, we demonstrate the possible uses of embeddings based clustering in different biological tasks. We explore two cases where we use the embeddings of KGE models to generate clusters of biological entities such as drugs and polypharmacy side-effects. We use visual clustering as an example to demonstrate cluster separation on a 2D space. However, in real scenarios, clustering algorithms utilise the full dimensionality of embedding vectors to build richer semantics of outcome clusters. Fig 5 shows two scatter plots of the embeddings of drugs from the DrugBank_FDA dataset and the polypharmacy side-effects reduced to a 2D space. We reduced the

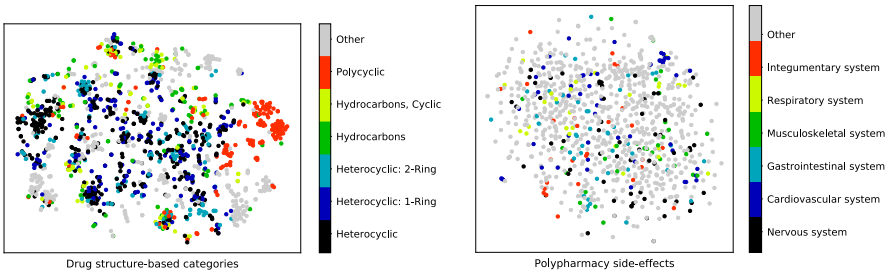

Figure 5: Three similarity matrices that denotes the Drug-drug similarities, motif-motif similarities and protein-protein similarities. The similarity values are generated by computing the cosine similarity between the embeddings of the pairs of compared entities. All the embeddings used to generated this figure are computed on the DrugBank_FDA datasets with the proteins associated to their PFam [90] motifs and protein families.

original embeddings using the T-SNE dimensionality reduction module [99] with the cosine distance configuration to reduce the embedding vectors to a 2D space.

The following examples examines two cases that differs in terms of the quality of generated clusters where we examine both drugs and polypharmacy side-effects according to different properties. In the first example (drug clustering), the generated embeddings is able to provide efficient clustering. On the other hand, in the second example, the polypharmacy side-effects, the learnt embeddings could not be separated into visible clusters according to the investigated property.

- **Clustering drugs** The left plot in Fig. 5 shows a scatter plot of the reduced embedding vectors of drugs coloured according to their chemical structure properties. The drugs are annotated with seven different chemical structure annotations: *Polycyclic*, *Hydrocarbons Cyclic*, *Hydrocarbons*, *Heterocyclic*, *Heterocyclic 1-Ring*, *Heterocyclic 2-Ring* and other chemicals. These annotations represent the six most frequent drug chemical structure category annotation extracted from the DrugBank database.

  We can see in the plot that the *Polycyclic* chemicals are located within a distinguishable cluster in the right side of the plot. The plot also shows that other types of *Hydrocarbons* and *Heterocyclic* chemicals form different micro-clusters in different locations in the plot.

  These different clusters can be used to represent a form of similarity between the different drugs. It can also be used to examine the relation between the embeddings as a representation with the original attributes of the examined drugs.

- **Clustering polypharmacy side-effects** The right plot in Fig. 5 shows a scatter plot of the reduced embedding vectors of polypharmacy side-effects. The plot polypharmacy side-effect points are coloured according to the human

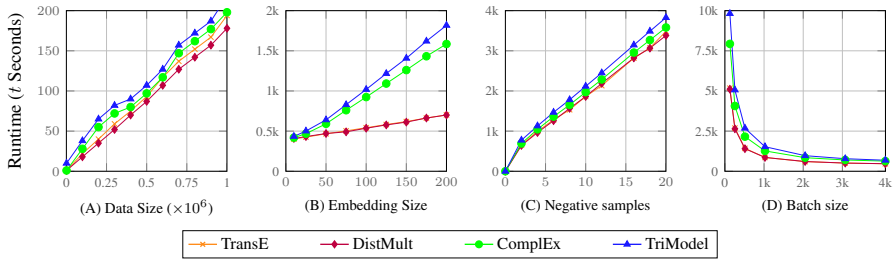

Figure 6: A set of line plots that describe the relation between the training runtime and the data size and configurable parameters of the TransE, DistMult, ComplEx and TriModel knowledge graph embedding models. The y-axis in all the plots represents the training time in seconds with different scales while the x-axis represents the data size and the models' parameters embedding size, negative samples and batch size respectively. The reported results are acquired by running the KGE models on the polypharmacy side-effects' full dataset ($\approx 4.5M$ instances).

body systems they affect. The plot includes a set of six categories of polypharmacy side-effects that represent six different human body systems *e.g.* nervous system.

Unlike the drug clusters illustrated in the left plot, the polypharmacy side-effects system-based categorisation does not yield obvious clusters. They, however, form tiny and scattered groups across the plot. This shows that the KGE models are unable to learn representations that can easily separate polypharmacy side-effects according to their associated body system.

5 Practical considerations for KGE models

In this section, we discuss different practical considerations related to the use of KGE models. We discuss their scalability on different experimental configurations, and we explore their different training and implementation strategies.

5.1 Scalability

Not only KGE models outperform other approaches in biological knowledge graphs completion tasks, but they also have better scalability compared to usual graph exploratory approaches. Often, complex biological systems are modelled as graphs where exploratory graph analytics methods are applied to perform different predictive tasks [5–7]. These models however suffer from limited scalability as they depend on graph traversal techniques that require complex training and predictions times [100, 101]. On the other hand, KGE models operate using linear time and space complexity [29, 59].

On the other hand, explanatory graph models use graph path searches which require higher time and space complexity [22]. For example, the DDR model [21]

is an exploratory graph drug-target predictor which uses graph random walks as features. A recent study [102] has shown that knowledge graph embedding models can outperform such models with higher scalability and better predictive accuracy. This is due to their linear time and space complexity procedures [29] compared to other exploratory models which use polynomial and exponential time and space procedures [23, 103].

In the following, we provide an empirical study of the scalability of KGE models in terms of different experimental configuration. We have studied the relation between the training runtime of KGE models and several training configuration parameters to examine their scalability capabilities. We have investigated the relation between the training runtime and the data size, embedding size, training negative samples and the training data batch size. We have performed our study on the polypharmacy side-effects data where the objective was to learn embeddings of drugs and polypharmacy side-effects.

Fig. 6 shows the outcome results of our study across the different investigated attributes. The plot "A" shows the relation between the training runtime and the size of the processed data. The plot shows that all the four investigated have a linear relation between their training runtime and the investigated data size. The plot also shows that the investigated models have a consistent growth in terms of their runtime across all the data sizes. The DistMult model consistency achieves the smallest runtime followed by the TransE, ComplEx and TriModel models respectively.

Plot "B" shows the relationship between the training runtime and the model embedding size. The plot shows that all the investigated models have a linear growth of their training runtime corresponding to the growth of the embeddings size. However, the growth rate of the TransE and DistMult models is considerably smaller than the growth of both the ComplEx and TriModel models. This occurs as both the TransE and DistMult models use a single vector to represent each of their embeddings while the ComplEx and TriModel models use two and three vectors respectively. Despite the better scalability of both the TransE and DistMult models, the ComplEx and TriModel models generally achieve better predictive accuracy than the TransE and DistMult models [64].

The plot "C" shows the relation between the runtime of KGE models and the number of negative samples they use during training. The plot shows that there is a positive linear correlation between training runtime and the number of negative samples–where all the KGE models have similar results across all the investigated sampling sizes. The TriModel, however, consistently have the highest runtime compared to other models.

Plot "D" shows the effects of the size of the batch on the training runtime. The plot shows an exponential decay of the training runtime with the linear growth of the data batch size. The KGE models process all the training data for each training iteration *i.e.* epoch, where the data is divided into batches for scalability and generalisation purposes. Therefore, the increase of the training data batch sizes lead to a decrease of the number of model executions for each training iteration. Despite the high scalability that can be achieved with large batch sizes, the best predictive accuracy is often achieved using small data batch sizes. Usually, the most efficient training data batch size is chosen during a hyper-parameter grid search along with other

parameters such as the embedding size and the number of negative samples.

5.2 Implementation and training strategies

Different implementations of KGE models are available online in different repositories as shown in Table. 2. The high scalability of KGE models allows them to be ported to both CPUs and GPUs where they can benefit from the high performance capabilities of GPU cores. They can also be implemented to operate in a multi-machine design, where they perform embedding training in a distributed fashion [104]. This configuration is better suited for processing knowledge graph of massive volumes that is hard to fit into one machine.

In this study, all our experiments are implemented in Python 3.5 using the Tensorflow library where we train our models on a single GPU card on one machine. We run our experiments on a Linux machine with an Intel(R) Core(TM) i7 processor, 32 GB RAM, and an nVidia Titan Xp GPU.

6 Opportunities and challenges

In this section, we discuss the challenges and opportunities related to the general and biological applications of KGE models. We begin by discussing the scope of input data for these models. We then discuss possible applications of KGE models in the biological domain. We conclude by discussing the limited interpretability of KGE models and other general limitations related to their biological applications.

6.1 Potential applications

KGE models can build efficient representations of biological data which is modelled as 3D tensors or knowledge graphs. This includes multiple types of biological data such as protein interactome and drug target interactions. In the following, we discuss examples of biological tasks and applications that can be performed using KGE models.

- **Modelling proteomics data.** KGE models can be used to model the different types of protein–protein interactions such as binding, phosphorylation, etc [105, 106]. This can be achieved by modelling these interactions as a knowledge graphs and applying the KGE models to learn the embeddings of the different proteins and interaction types. They can also be used to model the tissue context of interactions where different body tissues have different expression profiles of proteins, and these differences in expression affect the the proteins' interaction network. KGE can be used to model these interactions with their associated contexts as tensors [6].

  The biological activities of proteins also differ depending on their tissue context [86]. This type of information can easily be modelled using tensors where KGE models can be used to analyse the different functions of proteins depending on their tissue context [107].

- **Modelling genomics data.** Genomics data has been widely used to predict multiple gene associated biological entities such as gene–disease and gene–function associations [108, 109]. These approaches model the gene association in different ways including tensors and graph based representations [110]. KGE models can be easily utilised to process such data and provide efficient representations of genes and their associated biological objects. They can be further used to analyse and predict new disease–gene and gene–function associations.

- **Modelling pharmacological systems.** Information on pharmaceutical chemical substances is becoming widely available on different knowledge bases [46, 71]. This information includes the drug–drug and drug–protein interactome. In this context, KGE models can be a natural fit, where they can be used to model and extend the current pharmacological knowledge. They can also be used to model and predict both traditional and polypharmacy side-effects of drugs as shown in recent works [8, 111].

More details and discussion of the possible uses of KGE models and other general network embedding methods can be found in the study of Su et. al. [14] which discusses further potential uses of these methods in the biological domain.

6.2 Limitations of the KGE models

In the following, we discuss the limitations of the KGE models in both general and biological applications.

- **Lack of interpretability** In knowledge graph embedding models, the learning objective is to model nodes and edges of the graph using low-rank vector embeddings that preserve the graph's coherent structure. The embedding learning procedure operates mainly by transforming noise vectors to useful embeddings using gradient decent optimisation on a specific objective loss. Despite the high accuracy and scalability of this procedure, these models work as a black box and they are hard to interpret. Some approaches have suggested enhancing the interpretability of KGE models by using constraining training with a set of predefined rules such as type constraints [112], basic relation axioms [113], etc. These approaches thus enforce the KGE models to learn embeddings that can be partially interpretable by their employed constraints.

  In recent studies, researchers have also explored the interpretability of KGE models through new predictive approaches on top of the KGE models. For example, Gusmão et. al. [114] suggested the use of pedagogical approaches where they have used an alternative graphical predictive model, the SFE model [19], to link the learnt graph embeddings to the original knowledge graph. This approach was able to provide a new way for finding links between the embeddings and the original knowledge; however, the outcomes of these methods are still limited by the expressibility and feature coverage of the newly employed predictive models. The interpreting method in this context, also depends on

graph traversal methods which have limited scalability on large knowledge graphs [20].

- **Data quality** KGE models generate vector representations of biological entities according to their prior knowledge. Therefore, the quality of this knowledge affects the quality of the generated embeddings. For example, there is a high variance in the available prior knowledge on proteins where well studied proteins have significantly higher coverage in most databases [115]. This has a significant impact on quality of the less represented proteins as KGE models will be biased towards more studied proteins (*i.e.* highly covered proteins).

  In recent years, multiple works have explored the quality of currently available knowledge graphs [116] and the effect of low quality graphs on embedding models [117]. These works have shown that the accuracy KGE predictions degrade as sparsity and unreliability increase [117].

  This issue can be addressed by extending the available knowledge graph facts through merging knowledge bases of similar content. For example, drug target prediction using KGE models can be enhanced by extending the knowledge of protein–drug interactions by extra information such as protein-protein interactions and drug properties [102].

- **Knowledge evolution** Biological knowledge evolves everyday, where new chemicals and drugs are introduced and different associations between biological entities are discovered. However, KGE models in this context, are unable to encode the newly introduced entities. This results from their dependence on prior knowledge instead of the structural informations of proteins and chemical substances.

  This issue can be addressed by combining knowledge graph embedding scoring procedure with other sequence and structured based scoring mechanisms. This can allow informed prediction on new unknown objects. However, such a strategy will affect the scalability of predictions due to the newly introduced sequence and structure based features.

- **Hyper-parameter sensitivity** The outcome predictive accuracy of KGE embeddings is sensitive to their hyper-parameters [118]. Therefore, minor changes in these parameters can have significant effects on the outcome predictive accuracy of KGE models. The process of finding the optimal parameters of KGE models is traditionally achieved through an exhausting brute-force parameter search. As a result, their training may require rather time-consuming grid search procedure to find the right parameters for each new dataset.

  In this regard, new strategies for hyper parameter tuning such as differential evolution [119], random searches [120] and Bayesian hyper parameter optimisation [121]. These strategies can yield a more informed parameter search results with less running time.

- **Reflecting complex semantics of biological data in models based on knowledge graphs** Knowledge graph embedding methods are powerful in encoding

direct links between entities, however, they have limited ability in encoding simple indirect semantics such as types at different abstraction levels (*i.e.* taxonomies). For example, a KGE model can be very useful in encoding networks of interconnecting proteins which are modelled using direct relations. However, it has limited ability in encoding compound, multi-level relationships such as protein involvement in diseases due to their involvement in pathways that cause this disease. Such compound relationships that could be used for modelling complex biological knowledge are notoriously hard to reflect in KGE models [122]. However, the KGE models do have some limited ability to encode for instance type constrains [123], basic triangular rules [122] or cardinality constraints [124]. This could be used for modelling complex semantic features reflecting biological knowledge in future works. One has to bear in mind, though, that the designs of these semantics-enhanced KGE models typically depends on an extra computational routines to regularise the learning process which affects their scalability.

In their study, Su et. al. [14] have also discussed further general limitations of network embedding methods, and the effects and consequences of such limitations on the use of network embedding methods in the biological domain.

7 Conclusions

In this study, we discussed knowledge graph embedding (KGE) models and their biological applications. We presented two biological case studies, predicting drug targets and predicting polypharmacy side-effects, to demonstrate the predictive and analytical capabilities of KGE models. We demonstrated by computational experimental evaluation that KGE models outperform state-of-the-art approaches in solving the two studied problems on standard benchmarks. We also demonstrated the analytical capabilities of KGE such as clustering and measuring concept similarities. In this regard, we demonstrated KGE models' abilities to learn efficient similarities between different biological entities such as drugs and proteins. We also showed that the KGE models can efficiently be used as clustering methods for biological entities.

Furthermore, we discussed different practical considerations regarding the scalability and training strategies of KGE models. We also discussed the potential applications of KGE models in the biological domain. We finally discussed the challenges and limitations which face KGE models where we explored both their general limitations and the challenges that face them in the biological domain. In conclusion, we believe that the presented study can be a solid stepping stone towards many promising applications of the emergent KGE technology in the field of computational biology.

References

[1] Jonathan D. Cohen and David Servan-Schreiber. Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychological review*, 99 1:45–77, 1992.

[2] Jean-Francois Gibrat, Thomas Madej, and Stephen H. Bryant. Surprising similarities in structure comparison. *Current opinion in structural biology*, 6 3:377–85, 1996.

[3] Albert-László Barabási and Zoltán N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101–113, 2004.

[4] Réka Albert. Scale-free networks in cell biology. *Journal of cell science*, 118 Pt 21:4947–57, 2005.

[5] Vuk Janjic and Natasa Przulj. Biological function through network topology: a survey of the human diseasome. *Briefings in functional genomics*, 11 6:522–32, 2012.

[6] Emir Muñoz, Vít Novácek, and Pierre-Yves Vandenbussche. Facilitating prediction of adverse drug reactions by using knowledge graphs and multi-label learning models. *Briefings in bioinformatics*, 2017.

[7] Rawan S Olayan, Haitham Ashoor, and Vladimir B Bajic. Ddr: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. *Bioinformatics*, 34(7):1164–1173, 2017.

[8] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. In *Bioinformatics*, 2018.

[9] Sameh K. Mohamed, Vít Nováček, and Aayah Nounu. Drug target discovery using knowledge graph embeddings. In *Proceedings of the 34rd Annual ACM Symposium on Applied Computing*, SAC '19, pages 11–18, 2019.

[10] The UniProt Consortium. Uniprot: the universal protein knowledgebase. In *Nucleic Acids Research*, 2017.

[11] The Gene Ontology Consortium. The gene ontology resource: 20 years and still going strong. In *Nucleic Acids Research*, 2019.

[12] Michel Dumontier, Alison Callahan, and Jose Cruz-Toledo et. al. Bio2rdf release 3: A larger, more connected network of linked data for the life sciences. In *Proceedings of the ISWC 2014 Posters & Demonstrations.*, pages 401–404, 2014.

[13] Mona Alshahrani, Mohammed Asif Khan, and Omar Maddouri at. al. Neuro-symbolic representation learning on biological knowledge graphs. In *Bioinformatics*, 2017.

[14] Chang Su, Jie Tong, Yongjun Zhu, Peng Cui, and Fei Wang. Network embedding in biomedical data science. *Briefings in bioinformatics*, 2018.

[15] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016.

[16] Ni Lao, Tom Michael Mitchell, and William W. Cohen. Random walk inference and learning in a large scale knowledge base. In *EMNLP*, 2011.

[17] Bin Xu, Jihong Guan, Yang Wang, and Zewei Wang. Essential protein detection by random walk on weighted protein-protein interaction networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16:377–387, 2017.

[18] Karthik Raman. Construction and analysis of proteinâĂŞprotein interaction networks. In *Automated experimentation*, 2010.

[19] Matt Gardner and Tom M. Mitchell. Efficient and expressive knowledge base completion using subgraph feature extraction. In *EMNLP*, pages 1488–1498. The Association for Computational Linguistics, 2015.

[20] Sameh K. Mohamed, Vít Nováček, and Pierre-Yves Vandenbussche. Knowledge base completion using distinct subgraph paths. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC '18, pages 1992–1999, 2018.

[21] Rawan S. Olayan, Haitham Ashoor, and Vladimir B. Bajic. DDR: efficient computational method to predict drug-target interactions using graph mining and machine learning approaches. *Bioinformatics*, 34(7):1164–1173, 2018.

[22] Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66. ACL, 2015.

[23] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016.

[24] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.*, 29(12):2724–2743, 2017.

[25] Timothée Lacroix, Nicolas Usunier, and Guillaume Obozinski. Canonical tensor decomposition for knowledge base completion. In *ICML*, volume 80 of *JMLR Workshop and Conference Proceedings*, pages 2869–2878. JMLR.org, 2018.

[26] Antoine Bordes, Nicolas Usunier, and Alberto García-Durán et. al. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 2787–2795, 2013.

[27] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *ICML*, pages 809–816. Omnipress, 2011.

[28] Bishan Yang, Wen-tau Yih, and Xiaodong He et. al. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*, 2015.

[29] Théo Trouillon, Johannes Welbl, and Sebastian Riedel et. al. Complex embeddings for simple link prediction. In *ICML*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 2071–2080. JMLR.org, 2016.

[30] Tim Dettmers, Minervini Pasquale, Stenetorp Pontus, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, February 2018.

[31] Marinka Zitnik and Blaz Zupan. Collective pairwise classification for multi-way analysis of disease and drug data. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 21:81–92, 2016.

[32] Ibrahim Abdelaziz, Achille Fokoue, and Oktie Hassanzadeh et. al. Large-scale structural and textual similarity-based mining of knowledge graph to predict drug-drug interactions. *J. Web Semant.*, 44:104–117, 2017.

[33] Richard Qian. Understand your world with bing, 2013. Bing Blogs.

[34] David A. Ferrucci, Eric W. Brown, and Jennifer Chu-Carroll et. al. Building watson: An overview of the deepqa project. *AI Magazine*, 31(3):59–79, 2010.

[35] Tom M. Mitchell, William W. Cohen, and Estevam R. Hruschka Jr. et. al. Never-ending learning. In *AAAI*, pages 2302–2310. AAAI Press, 2015.

[36] George A. Miller. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41, November 1995.

[37] Yongjun Zhu, Olivier Elemento, Jyotishman Pathak, and Fei Wang. Drug knowledge bases and their applications in biomedical informatics research. *Briefings in bioinformatics*, 2018.

[38] Alan R. Aronson, James G. Mork, Clifford W. Gay, Susanne M. Humphrey, and Willie J. Rogers. The nlm indexing initiative's medical text indexer. *Studies in health technology and informatics*, 107 Pt 1:268–72, 2004.

[39] Melissa J. Landrum, Jennifer M. Lee, and George R. Riley et. al. Clinvar: public archive of relationships among sequence variation and human phenotype. In *Nucleic Acids Research*, 2014.

[40] Minoru Kanehisa, Miho Furumichi, and Mao Tanabe et. al. Kegg: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361, 2017.

[41] Sandra E. Orchard, Mais G. Ammari, and Bruno Aranda et. al. The mintact project intact as a common curation platform for 11 molecular interaction databases. In *Nucleic Acids Research*, 2014.

[42] Antonio Fabregat, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, Bijay Jassal, Florian Korninger, Bruce May, Marija Milacic, Corina Duenas, Karen Rothfels, Cristoffer Sevilla,

Veronica Shamovsky, Solomon Shorser, Thawfeek M. Varusai, Guilherme Viteri, Joel Weiser, Guanming Wu, Lincoln Stein, Henning Hermjakob, and Peter D'Eustachio. The reactome pathway knowledgebase. In *Nucleic Acids Research*, 2018.

[43] Minoru Kanehisa, Yoko Sato, and Masayuki Kawashima et. al. Kegg as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1):D457–D462, 2016.

[44] David S. Wishart, Craig Knox, and An Chi Guo et. al. Drugbank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research*, 36:D901–D906, 2008.

[45] Carolyn J. Mattingly, Glenn T Colby, John N. Forrest, and James Boyer. The comparative toxicogenomics database (ctd). *Environmental Health Perspectives*, 111:793 – 795, 2003.

[46] Anna Gaulton, Anne Hersey, and Michal Nowotka et. al. The chembl database in 2017. In *Nucleic Acids Research*, 2017.

[47] Michael Kuhn, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. The sider database of drugs and side effects. *Nucleic acids research*, 44 D1:D1075–9, 2016.

[48] Mathias Uhlén, Linn Fagerberg, and Björn Mikael Hallström et. al. Tissue-based map of the human proteome. *Science*, 347, 2015.

[49] Damian Szklarczyk, John H. Morris, and Helen Victoria Cook et. al. The string database in 2017: quality-controlled proteinâĂŞprotein association networks, made broadly accessible. In *Nucleic Acids Research*, 2017.

[50] Chris Stark, Bobby-Joe Breitkreutz, and Andrew Chatr aryamontri et. al. The biogrid interaction database: 2011 update. *Nucleic acids research*, 39 Database issue:D698–704, 2007.

[51] Alex L Mitchell, Teresa K Attwood, Patricia C Babbitt, and et. al. Interpro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, 47(D1):D351–D360, 2019.

[52] Micheal Hewett, Diane E. Oliver, Daniel L. Rubin, Katrina L. Easton, Joshua M. Stuart, Russ B. Altman, and Teri E. Klein. Pharmgkb: the pharmacogenetics knowledge base. *Nucleic acids research*, 30 1:163–5, 2002.

[53] X. Chen, Zhi Liang Ji, and Yu Zong Chen. Ttd: Therapeutic target database. *Nucleic acids research*, 30 1:412–5, 2002.

[54] Nikolai Hecker, Jessica Ahmed, and Joachim von Eichborn et. al. Supertarget goes quantitative: update on drugâĂŞtarget interactions. In *Nucleic Acids Research*, 2012.

[55] François Belleau, Marc-Alexandre Nolin, Nicole Tourigny, Philippe Rigault, and Jean Morissette. Bio2rdf: towards a mashup to build bioinformatics knowledge systems. *Journal of biomedical informatics*, 41(5):706–716, 2008.

[56] Christian Bizer and Richard Cyganiak. D2r server-publishing relational databases on the semantic web. In *Poster at the 5th international semantic web conference*, volume 175, 2006.

[57] Siham Amrouch and Sihem Mostefai. Survey on the literature of ontology mapping, alignment and merging. In *2012 International Conference on Information Technology and e-Services*, pages 1–5. IEEE, 2012.

[58] Axel-Cyrille Ngonga Ngomo and Sören Auer. LimesâĂŤa time-efficient approach for large-scale link discovery on the web of data. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[59] Sameh K. Mohamed, Emir Muñoz, Vít Nováček, and Pierre-Yves Vandenbussche. Loss functions in knowledge graph embedding models. In *DL4KGS@ESWC*, volume 2106 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2019.

[60] Antoine Bordes, Xavier Glorot, Jason Weston, and Yoshua Bengio. A semantic matching energy function for learning with multi-relational data - application to word-sense disambiguation. *Machine Learning*, 94(2):233–259, 2014.

[61] Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In *EMNLP*, 2016.

[62] Maximilian Nickel, Lorenzo Rosasco, and Tomaso A. Poggio. Holographic embeddings of knowledge graphs. In *AAAI*, pages 1955–1961. AAAI Press, 2016.

[63] Hanxiao Liu, Yuexin Wu, and Yiming Yang. Analogical inference for multi-relational embeddings. In *ICML*, 2017.

[64] Sameh K. Mohamed and Vít Novácek. Link prediction using multi part embeddings. In *ESWC*, volume 11503 of *Lecture Notes in Computer Science*, pages 240–254. Springer, 2019.

[65] Bryan Perozzi, Rami Al-Rfou', and Steven Skiena. Deepwalk: online learning of social representations. *ArXiv*, abs/1403.6652, 2014.

[66] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. *KDD: proceedings. International Conference on Knowledge Discovery & Data Mining*, 2016:855–864, 2016.

[67] Georg C Terstappen, Christina Schlüpen, Roberto Raggiaschi, and Giovanni Gaviraghi. Target deconvolution strategies in drug discovery. *Nature Reviews Drug Discovery*, 6(11):891, 2007.

[68] Lekha Sleno and Andrew Emili. Proteomic methods for drug target discovery. *Current opinion in chemical biology*, 12(1):46–54, 2008.

[69] Yoshihiro Yamanishi, Michihiro Araki, and Alex. Gutteridge et. al. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 24(13):i232–i240, 2008.

[70] Jian-Ping Mei, Chee-Keong Kwoh, and Peng Yang et. al. Drug–target interaction prediction by learning from local information and neighbors. *Bioinformatics*, 29(2):238–245, 2012.

[71] David S. Wishart, Craig Knox, and An Chi Guo et. al. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34:D668–D672, 2006.

[72] Feixiong Cheng, Yadi Zhou, and Weihua Li et. al. Prediction of chemical-protein interactions network with weighted network-based inference method. In *PloS one*, 2012.

[73] Feixiong Cheng, Chuang Liu, and Jing Jiang et. al. Prediction of Drug-Target Interactions and Drug Repositioning via Network-Based Inference. In *PLoS Computational Biology*, 2012.

[74] Ayeshah A Rosdah, Jessica K. Holien, and Lea MD Delbridge et. al. Mitochondrial fission–a drug target for cytoprotection or cytodestruction? *Pharmacology research & perspectives*, 4(3):e00235, 2016.

[75] Hui Liu, Jianjiang Sun, and Jihong Guan et. al. Improving compound–protein interaction prediction by building up highly credible negative samples. *Bioinformatics*, 31(12):i221–i229, 2015.

[76] André CA Nascimento, Ricardo BC Prudêncio, and Ivan G Costa. A multiple kernel learning algorithm for drug-target interaction prediction. *BMC bioinformatics*, 17(1):46, 2016.

[77] Ming Hao, Stephen H Bryant, and Yanli Wang. Predicting drug-target interactions by dual-network integrated logistic matrix factorization. *Scientific reports*, 7:40376, 2017.

[78] Joanne Bowes, Andrew J Brown, and Jacques Hamon et. al. Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. *Nature Reviews Drug discovery*, 11(12):909–922, 2012.

[79] Elizabeth D. Kantor, Colin D Rehm, and Jennifer S. Haas et. al. Trends in prescription drug use among adults in the united states from 1999-2012. *JAMA*, 314 17:1818–31, 2015.

[80] Nicholas P. Tatonetti, Patrick Ye, Roxana Daneshjou, and Russ B. Altman. Data-driven prediction of drug effects and interactions. *Science translational medicine*, 4 125:125ra31, 2012.

[81] Alberto García-Durán and Mathias Niepert. Kblrn: End-to-end learning of knowledge base representations with latent, relational, and numerical features. In *UAI*, 2018.

[82] Linn Fagerberg, Björn M Hallström, Per Oksvold, Caroline Kampf, Dijana Djureinovic, Jacob Odeberg, Masato Habuka, Simin Tahmasebpoor, Angelika Danielsson, Karolina Edlund, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics*, 13(2):397–406, 2014.

[83] Casey S Greene, Arjun Krishnan, Aaron K Wong, Emanuela Ricciotti, Rene A Zelaya, Daniel S Himmelstein, Ran Zhang, Boris M Hartmann, Elena Zaslavsky, Stuart C Sealfon, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nature genetics*, 47(6):569, 2015.

[84] Vivette D D'Agati. The spectrum of focal segmental glomerulosclerosis: new insights. *Current opinion in nephrology and hypertension*, 17(3):271–281, 2008.

[85] James J Cai and Dmitri A Petrov. Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes. *Genome biology and evolution*, 2:393–409, 2010.

[86] Marinka Zitnik and Jure Leskovec. Predicting multicellular function through multi-layer tissue networks. In *Bioinformatics*, 2017.

[87] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *WWW*, 2015.

[88] David Warde-Farley, Sylva L. Donaldson, Ovi Comes, Khalid Zuberi, Rashad Badrawi, Pauline Chao, Max Franz, Chris Grouios, Farzana Kazi, Christian Tannus Lopes, Anson Maitland, Sara Mostafavi, Jason Montojo, Quentin Shao, George Wright, Gary D. Bader, and Quaid Morris. The genemania prediction server: biological network integration for gene prioritization and predicting gene function. In *Nucleic Acids Research*, 2010.

[89] Hansaim Lim, Paul Gray, Lei Xie, and Aleksandar Poleksic. Improved genome-scale multi-target virtual screening via a novel collaborative filtering approach to cold-start problem. *Scientific reports*, 6:38860, 2016.

[90] Alex Bateman, Lachlan James M. Coin, and Richard Durbin et. al. The pfam protein families database. *Nucleic acids research*, 28 1:263–6, 2000.

[91] Brandon Malone, Alberto García-Durán, and Mathias Niepert. Knowledge graph completion to predict polypharmacy side effects. In *DILS*, 2018.

[92] Evangelos E. Papalexakis, Christos Faloutsos, and Nikos D. Sidiropoulos. Tensors for data mining and data fusion: Models, applications, and scalable algorithms. *ACM TIST*, 8:16:1–16:44, 2016.

[93] Werner L. Lipschitz, Zareh Hadidian, and Andrew Kerpcsar. Bioassay of diuretics. 1943.

[94] Johann Emanuel Pohl, Herbert F. Thurston, and John D. Swales. The antidiuretic action of diazoxide. 1972.

[95] Joris C Verster and Edmund R. Volkerts. Clinical pharmacology, clinical efficacy, and behavioral toxicity of alprazolam: a review of the literature. *CNS drug reviews*, 10 1:45–76, 2004.

[96] John P. Overington, Bissan Al-Lazikani, and Andrew L. Hopkins. How many drug targets are there? *Nature Reviews Drug Discovery*, 5:993–996, 2006.

[97] Yasuji Minoda and Evan D Kharasch. Halothane-dependent lipid peroxidation in human liver microsomes is catalyzed by cytochrome p4502a6 (cyp2a6). *Anesthesiology*, 95 2:509–14, 2001.

[98] Krisna Rungruangsak-Torrissen, Charles Garvie Carter, and Anne Sundby et. al. Maintenance ration, protein synthesis capacity, plasma insulin and growth of atlantic salmon (salmo salar l.) with genetically different trypsin isozymes. *Fish Physiology and Biochemistry*, 21:223–233, 1999.

[99] Laurens van der Maaten. Accelerating t-sne using tree-based algorithms. *Journal of Machine Learning Research*, 15:3221–3245, 2014.

[100] To-Yat Cheung. Graph traversal techniques and the maximum flow problem in distributed computation. *IEEE Transactions on Software Engineering*, (4):504–512, 1983.

[101] Pierre Fraigniaud, Leszek Gasieniec, Dariusz R Kowalski, and Andrzej Pelc. Collective tree exploration. *Networks: An International Journal*, 48(3):166–177, 2006.

[102] Sameh K. Mohamed, Vít Novácek, and Aayah Nounu. Discovering protein drug targets using knowledge graph embeddings. *Bioinformatics*, 2019.

[103] Sameh K. Mohamed, Emir Muñoz, Vít Novácek, and Pierre-Yves Vandenbussche. Identifying equivalent relation paths in knowledge graphs. In *LDK*, 2017.

[104] Adam Lerer, Ledell Wu, and Jiajun Shen et. al. Pytorch-biggraph: A large-scale graph embedding system. In *The 2nd SysML Conference*, 2019.

[105] Nurcan Tuncbag, Gozde Kar, Ozlem Keskin, Attila Gürsoy, and Ruth Nussinov. A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Briefings in bioinformatics*, 10 3:217–32, 2008.

[106] Jian Zhang and Lukasz A. Kurgan. Review and comparative assessment of sequenceâĂŘbased predictors of proteinâĂŘbinding residues. *Briefings in Bioinformatics*, 19:821âĂŞ837, 2018.

[107] Sameh K. Mohamed. Predicting tissue-specific protein functions using multi-part tensor decomposition. *Inf. Sci.*, 508:343–357, 2020.

[108] Michael J Bamshad, Sarah Boonhsi Ng, and Abigail W. Bigham et. al. Exome sequencing as a tool for mendelian disease gene discovery. *Nature Reviews Genetics*, 12:745–755, 2011.

[109] Xiangxiang Zeng, Ningxiang Ding, Alfonso Rodríguez-Patón, and Quan Zou. Probability-based collaborative filtering model for predicting geneâĂŞdisease associations. In *BMC Medical Genomics*, 2017.

[110] Anna Bauer-Mehren, Markus Bundschus, and Michael Rautschka et. al. Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. In *PloS one*, 2011.

[111] Emir Muñoz, Vít Novácek, and Pierre-Yves Vandenbussche. Using drug similarities for discovery of possible adverse reactions. In *AMIA 2016, American Medical Informatics Association Annual Symposium, Chicago, IL, USA, November 12-16, 2016*. AMIA, 2016.

[112] Denis Krompass, Stephan Baier, and Volker Tresp. Type-constrained representation learning in knowledge graphs. In *International Semantic Web Conference*, 2015.

[113] Pasquale Minervini, Luca Costabello, and Emir Muñoz et. al. Regularizing knowledge graph embeddings via equivalence and inversion axioms. In *ECML/PKDD*, 2017.

[114] Arthur Colombini Gusmão, Alvaro Henrique Chaim Correia, Glauber De Bona, and Fábio Gagliardi Cozman. Interpreting embedding models of knowledge bases: A pedagogical approach. In *Proceddings of WHI*, 2018.

[115] The Uniprot Consortium. Uniprot: a hub for protein information. In *Nucleic Acids Research*, 2015.

[116] Michael Färber, Frederic Bartscherer, Carsten Menne, and Achim Rettinger. Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago. *Semantic Web*, 9:77–129, 2017.

[117] Jay Pujara, Eriq Augustine, and Lise Getoor. Sparsity and noise: Where knowledge graph embeddings fall short. In *EMNLP*, 2017.

[118] Rudolf Kadlec, Ondrej Bajgar, and Jan Kleindienst. Knowledge base completion: Baselines strike back. In *Rep4NLP@ACL*, pages 69–74. Association for Computational Linguistics, 2017.

[119] Wei Fu, Vivek Nair, and Tim Menzies. Why is differential evolution better than grid search for tuning defect predictors? *ArXiv*, abs/1609.02613, 2016.

[120] Francisco J. Solis and Roger J.-B. Wets. Minimization by random search techniques. *Math. Oper. Res.*, 6:19–30, 1981.

[121] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. In *NIPS*, 2012.

[122] Leon Weber, Pasquale Minervini, Jannes Münchmeyer, Ulf Leser, and Tim Rocktäschel. Nlprolog: Reasoning with weak unification for question answering in natural language. In *ACL (1)*, pages 6151–6161. Association for Computational Linguistics, 2019.

[123] Pasquale Minervini, Luca Costabello, Emir Muñoz, Vít Novácek, and Pierre-Yves Vandenbussche. Regularizing knowledge graph embeddings via equivalence and inversion axioms. In *ECML/PKDD (1)*, volume 10534 of *Lecture Notes in Computer Science*, pages 668–683. Springer, 2017.

[124] Emir Muñoz, Pasquale Minervini, and Matthias Nickles. Embedding cardinality constraints in neural link predictors. In *SAC*, pages 2243–2250. ACM, 2019.